

基于增量学习的链路质量估计方法研究

靖浩翔, 施伟斌, 黄高远, 王振羽

上海理工大学光电信息与计算机工程学院, 上海

收稿日期: 2025年4月2日; 录用日期: 2025年5月2日; 发布日期: 2025年5月12日

摘要

链路质量估计是无线传感器网络建立多跳传输路径的基础, 受干扰、噪声和多径效应等因素影响, 采用离线训练的方法很难建立适用于各种条件的链路质量估计模型。文章提出一种基于增量学习的链路质量估计方法LEIL, 以在线获取的链路质量指标LQI和信噪比SNR作为输入特征, PRR作为预测目标, 在线建立链路估计模型。以EFD为基础, 采用均方误差MSE作为纯度指标, 实现了决策回归树的增量生成。实验结果显示, LEIL具有较高的预测精度, 不同链路质量条件下平均的MAE为0.0528。与现有方法相比, LEIL可以自适应环境, 无需重新建立训练数据集, 减小了建立链路估计模型的成本, 适用于各种工作条件的无线传感器网络。

关键词

增量学习, 决策树回归, 无线传感器网络, 链路质量估计

Research on Link Quality Estimation Method Based on Incremental Learning

Haoliang Jing, Weibin Shi, Gaoyuan Huang, Zhenyu Wang

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Apr. 2nd, 2025; accepted: May 2nd, 2025; published: May 12th, 2025

Abstract

Link quality estimation is the foundation for establishing multi-hop transmission paths in wireless sensor networks. Due to factors such as interference, noise, and multipath effects, it is difficult to establish a link quality estimation model suitable for various conditions using offline training methods. This paper proposes an incremental learning-based link quality estimation method LEIL, which takes the online-obtained link quality index LQI and signal-to-noise ratio SNR as input features, PRR

as the prediction target and establishes a link estimation model online. Based on EFDT and using mean square error (MSE) as the purity indicator, an incremental generation of decision regression trees has been achieved. The experimental results show that LEIL has high prediction accuracy, with an average MAE of 0.0528 under different link quality conditions. Compared with existing methods, LEIL can adapt to the environment without rebuilding the training dataset, reducing the cost of establishing a link estimation model, and is suitable for wireless sensor networks under various working conditions.

Keywords

Incremental Learning, Decision Tree Regression, Wireless Sensor Network, Link Quality Estimation

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

无线传感器网络(Wireless Sensor Networks, WSN)是一种由大量小型传感器节点组成的自组织网络,广泛应用于环境监测[1][2]、智能电网[3][4]、军事侦察[5]、医疗保健[6]和智能家居[7]等多个领域。在动态和复杂的环境中,传感器节点的通信质量可能受到多种因素的影响,包括多径效应、环境中其他网络的干扰以及硬件噪声等[8]。准确高效地估计链路质量对于提高数据传输效率、延长网络寿命和增强网络可靠性至关重要。

目前关于链路质量的估计方法往往采用采集数据集、离线训练模型、模型部署并预测的方式。然而,如果采集数据集所处的环境与实际部署模型的环境不一致,链路质量的预测精度可能有所下降。此外,传感器节点采集的物理层参数可能因不同生产厂家、计算标准不一导致用数据集离线训练的模型无法适配所有环境。

针对以上问题,本文利用 EFDT 决策回归树算法,采用在线学习的方法对链路质量进行预测。该方法将 SNR 和 LQI 作为模型的输入特征,与离线学习的决策回归树相比,该方法实时性高、泛化性强、能适应不同的环境。实验结果表明,本文提出的模型具有较高的预测精度,可以应用于无线传感器网络链路质量估计。

2. 相关研究

自从决策树算法提出以来,因其能适应复杂数据结构、支持多变量、模型具有较高可解释性等特点,受到学术界的广泛关注和持续研究。随着互联网和物联网技术的快速发展,众多组织面临着高速流入、持续增长的数据流。然而,传统决策树方法(如 ID3、C4.5)依赖于在内存中存储所有训练样本,这使得模型无法处理超出内存容量的数据或者大量数据未被利用,最终导致模型过度简单,从而引发欠拟合问题。此外,如果数据到达超过模型处理数据的速度,那么未处理的数据无限累积,存储和管理成本随之剧增。

针对以上挑战,在线学习方法提供了一种有效的解决方案。在此背景下专家学者针对传统决策树模型改进得到在线决策树算法。文献[9]提出了基于 Very Fast Decision Tree (VFDT)方法扩展得到的 VFDTc 算法。该方法使用动态二元树结构来维护连续变量的统计信息,通过遍历二元树结构快速找到最优分割点,从而避免了预先排序或者存储所有值,提高了算法的搜索效率。其次,该方法在叶节点采用朴素贝

叶斯分类器使得在处理小样本或者树分裂早期时的准确率优于多数类策略, 模型在 Waveform 数据集上实验误差降低了 5%~8%。

文献[10]针对特征候选分割点过多的问题将数值范围划分为等宽区间, 减少 39% 的连续特征处理时间, 并降低了内存需求。此外, 该文献修改了剪枝策略: 基于当前样本估计增益, 若某区间增益上限低于当前最佳增益则剪枝, 再通过反向检查避免误剪枝。实验显示这种剪枝策略没有显著损失精度。

文献[11]提出了一种基于 McDiarmid 不等式的决策树算法 McDiarmid Tree, 以解决传统的 Hoeffding Tree (HT) 算法的理论缺陷。作者推导出处理信息增益和基尼系数时的边界条件。此外, 该文献引入参数 τ , 当边界 $\epsilon < \tau$ 时强制分裂, 由此提升了构建决策树模型的效率。

相比之前提出的在线学习决策树算法, HT 更加接近离线学习中的决策树[12], 并且过去的算法不适合有噪声的数据流或者计算成本过高的场景[13]。文献[9]-[11]都是对 HT 算法的改进, 但是没有对分裂原理进行根本扩展。

文献[14]通过改进 VFDT 算法提出了 CVFDT 算法。该文献新增了滑动窗口机制和子树替代机制, 维护一个固定大小的窗口。当新样本被输入时, 更新相关节点的统计信息并移动滑动窗口, 移除旧数据的统计量, 根据窗口内的数据更新模型。随着时间或环境变化, 当模型检测某节点的分裂属性不再是最优的, 就基于最新数据生成新的子树替换原有的子树。滑动窗口机制使得更新每个样本的时间复杂度为 $O(1)$, 优于重新训练整个窗口的时间复杂度, 有效提升了建立决策树的效率。在概念漂移场景下, CVFDT 比 VFDT 的准确率提升了约 10%, 且节点数更少。

文献[15]改进了传统 HT 在线学习方法统计效率低以及无法修正分裂错误的问题, 提出 Hoeffding Anytime Tree (HATT), 该方法增加了及时分裂和分裂后修正策略, 从而显著增加建树效率。但是 HATT 仅讨论并实现分类问题的模型建立, 而没有谈回归问题。

针对无线传感器网络的链路质量估计也存在其他在线学习方法。文献[16]提出了 LEAS (Link Estimation with Asynchronous Samples) 的链路质量估计方法。通过节点周期性地采集噪声强度 (INS) 和干扰, 结合 RSSI 计算信噪比 SINR。使用简化的包成功率 (PSR)-SINR 线性映射模型, 利用 SINR 计算 PSR 的瞬时估计值, 并借助滑动窗口和 EWMA 平滑滤波。该方法有效降低了预测误差且无需离线训练模型, 但是其依赖于对 SINR 测量的准确性, 在存在 Wi-Fi、蓝牙干扰, 干扰源不可控的场景下, 该方法可能导致测量误差影响预测精度。

文献[17]针对 WSN 设备提出了一种基于蒙特卡洛模拟与非线性拟合的实时链路质量估计方法。该方法将符号错误率 (SER) 作为链路质量评价指标, 通过传感器采集的物理层参数 SNR 计算得到芯片错误率 (CER)。在离线训练中, 通过蒙特卡洛仿真生成实验数据, 使用非线性回归拟合建立 CER-SER 模型。在线预测中, 根据离线训练建立的模型实现实时计算。该方法计算复杂度低, 并且在典型噪声范围内表现稳定, 但是由于其依赖离线训练模拟数据, 所以若噪声超出模拟范围, 模型可能失效。其次, 模型参数可能因不同的硬件芯片差异需要重新校准, 泛化性受到限制。

文献[18]提出了一种针对工业无线传感器网络的实时链路质量估计方法。该方法在传感器网络中部署专属 LQE 节点, 通过独立射频模块避免占用传感器节点的通信资源, 持续监听接收节点的数据包, 获取 RSSI 和重复包数量。基于接收包的 RSSI 值, 通过离线训练的三次多项式回归模型实时估计前向 PRR。根据无数据接收时的 RSSI 值, 判断信道被外部干扰 (如 Wi-Fi、蓝牙) 占用的概率。统计重复包的数量并结合历史数据训练的模型估算后向 PRR。最后, 将以上三者通过 EWMA 平滑输出的加权结果作为总体链路质量的指标。该方法相比于 Opt-FLQE 方法降低了 20% 的重发数据。根据工业环境实测数据建模, 其适用于存在金属反射和多径效应的工业环境。

文献[19]提出了一种通过实时学习多个预测策略的表现, 动态调整权重以实现最小化预测误差的链

路质量动态预测方法。传感器节点获取实时 RSSI 和期望传输次数(EXT)并预测下一时刻的链路质量。同时,构建反馈机制,将真实的链路质量反馈给预测器。预测策略分为均值策略和保守策略,前者是利用滑动窗口内链路质量的均值估计链路质量,后者采用加权求和链路质量的均值和最小值以降低突发干扰影响。预测器根据学习率和损失函数调整每个策略的权重。实验结果显示,该方法在动态工业环境中相比于不实时更新模型参数的方法,预测误差降低 30%~50%。

文献[20]提出的在线学习方法与文献[19]类似,不同点在于将保守策略改为平滑因子滤波策略。该方法利用平滑因子调节历史数据和当前数据对策略输出结果的影响。此外,该方法直接选择历史累积损失最小的预测策略作为输出结果。该综合策略的方法在数据波动阶段比仅使用滑动窗口平均的 MSE 降低了约 30%, MAE 降低了 25%。采用直接选取最优策略的方法预测虽然节约了计算资源,但是在动态场景下,策略切换延迟导致的误差比窗口移动平均高 15%~20%。

以往的链路质量估计在线学习方法大致分为三类:第一类是采取离线训练映射模型,部署到节点上在线预测,但这会导致部署在非离线训练的特殊环境下模型可能会失效;第二类是额外部署链路质量估计节点,但这增加了成本;第三类是根据在线采集的实际数据和损失函数实时更新不同策略权重并进行预测,但这种方法往往增加了很多计算成本,且在环境动态场景下可能表现不佳。

3. EFDTR 决策回归树模型的建立

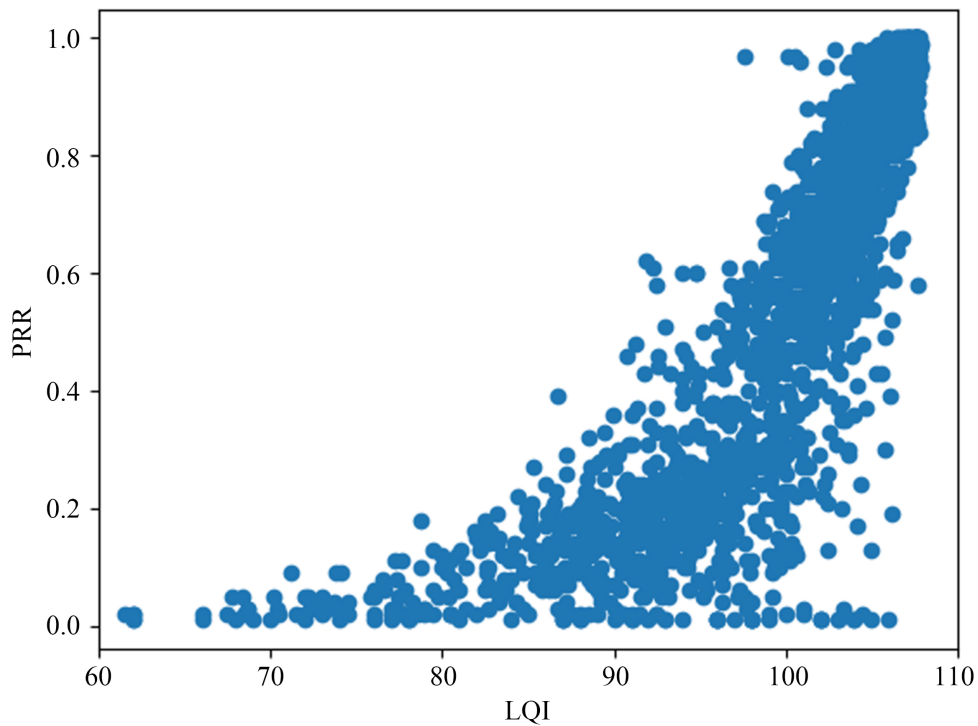


Figure 1. Scatter plot of PRR and LQI in experimental environment 1

图 1. 实验环境 1 中 PRR 与 LQI 的散点图

离线学习的方法利用特定环境下采集的数据集训练机器学习模型,然后部署到节点上完成在线预测。这种方法由于环境变化,物理层参数与 PRR 的映射关系也可能发生变化从而导致预测误差增大。由图 1 和图 2 所示,环境 1 和环境 2 中 PRR 和 LQI 的散点图分布不同,如果使用实验环境 1 中采集的数据集训练决策回归树模型并在实验环境 2 下进行预测,可能导致较大的预测误差。其次,在实际部

署的 WSN 中, 传感器节点的型号可能存在差异, 各个生产厂家对链路质量指标 LQI 的评定标准也可能有所不同。因此, 使用不同生产厂家的传感器节点采集数据得到的 PRR 和物理层参数之间的映射关系也可能不同, 使模型在不同设备中难以适用。由于传统决策树回归树的各种局限, 就需要一种泛化性强、能适应环境变化和不同物理器件且精度较高的机器学习模型, 在线学习的决策树模型可以满足以上要求。

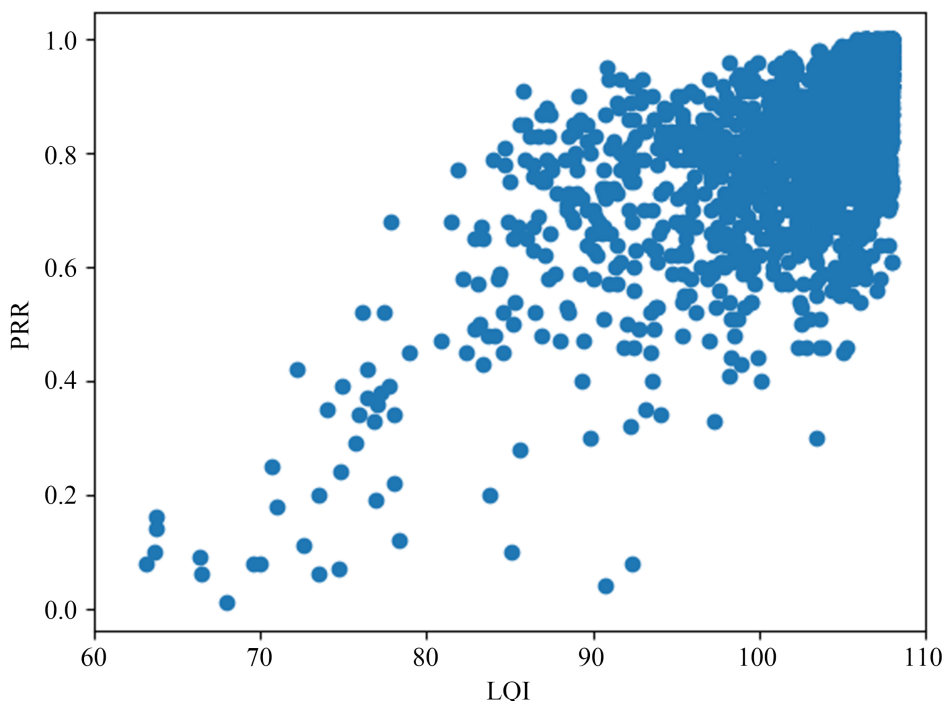


Figure 2. Scatter plot of PRR and LQI in experimental environment 2

图 2. 实验环境 2 中 PRR 与 LQI 的散点图

4. EFDT 算法原理

Domingos 和 Hulten 提出了第一个增量构建决策树的算法——快速决策树(VFDT) [21]。VFDT 也被称为 Hoeffding Tree (HT)算法, HT 的每个叶子节点会记录流入此节点样本的统计信息。其在训练初始时为单叶子节点, 也是树的根节点。随着实时数据更新, 新的数据流入到根节点并积累足够样本后, 根节点根据决策树的启发式度量遍历所有特征中的所有特征值, 寻找最优分裂。再根据霍夫丁边界判断是否满足分裂条件。若分裂条件达成, 就依据最优分裂将样本分成两个集合, 形成左右两个叶子节点, 否则继续积累样本。随着样本的持续更新, 新流入的每个样本依据分裂节点的分裂标准从根节点向下传递至叶子节点, 叶子更新其统计信息。当新形成的两个叶子节点累积到足够样本后, 再次启用启发式度量机制。按照以上步骤不断迭代从而形成一个处理数据流的增量决策树模型。

霍夫丁不等式是一种概率不等式, 它为独立、有界的随机变量之和偏离其期望值的概率提供了一个上限。根据霍夫丁不等式, 如果 n 个独立的随机变量 $X_1, X_2, \dots, X_n$, 每个变量的 X_i 都限制在某个取值范围 R 内[22]:

$$a_i \leq X_i \leq b_i \quad (1)$$

$$R = b_i - a_i \quad (2)$$

假定这些变量的均值为:

$$v_n = \frac{X_1 + X_2 + \cdots + X_n}{n} \quad (3)$$

这些变量的期望值为:

$$\mu = E(S_n) \quad (4)$$

其中 S_n 为随机变量的和, 那么这些变量的均值偏离期望值的概率被限定在一个范围内, 其满足双边不等式:

$$P(|v_n - \mu| \geq \varepsilon) \leq 2e^{-\frac{2n\varepsilon^2}{R^2}} \quad (5)$$

其中, ε 是偏差量。经转换后可以得到随机变量的实际均值在 $[v_n - \varepsilon, v_n + \varepsilon]$ 区间的置信度为 $1 - \delta$, ε 可以表示为:

$$\varepsilon = \sqrt{\frac{R^2 \ln \frac{1}{\delta}}{2n}} \quad (6)$$

HT 的核心策略是: 当候选分割属性间的基尼系数之差超过霍夫丁边界时, 选择最优属性进行节点分裂, 并且一旦完成了分割便不再做修改。这种保守策略虽然提高了计算效率, 却导致两个关键缺陷: 1) 早期学习阶段需积累大量样本才能达到分割阈值, 造成模型收敛延迟。2) 随着新数据到来, 原来的最优分割不再是最优时, 原始分割无法修正, 随着树深度增加, 模型与理想批量决策树的偏离逐渐变大。

基于此, EFDT 算法, 也称 HATT, 提出了改进思想。在 HATT 决策树中, 也采用霍夫丁不等式作为节点的分裂依据。EFDT 算法的核心是动态分割机制, 该机制通过两个决策阶段实现了决策树结构的实时校准与渐进优化: 在叶节点执行初始分割决策的即时分裂阶段, 以及在决策节点实施分割验证的持续优化阶段。在第一阶段, 样本数据流到达叶节点, 模型更新所有候选属性的统计量。如果累积到 $n_{\min}$ 个样本, 则计算各属性的基尼系数, 其中 $n_{\min}$ 是模型的分裂最小样本数。根据霍夫丁不等式, 当最优属性 X_a 与不分裂 $X_\emptyset$ 的基尼系数之差满足时, 节点立即执行分裂操作, 生成以 X_a 为分割属性的决策节点及左右两个子叶节点。其中, $R = \log C$, C 为样本目标类别数, δ 为置信度, n 为样本数。

$$\Delta G > \sqrt{\frac{R^2 \ln \frac{1}{\delta}}{2n}} \quad (7)$$

在第二阶段, 随着新的样本实时更新流入模型, 根据现有树结构流经决策节点到达叶子节点, 决策节点每流经 $n_{\min}$ 个样本后, 系统将重新评估当前分割属性 X_{current} 与新候选最优属性 X'_a 的基尼系数差:

$$\Delta G' = G(X_{\text{current}}) - G(X'_a) \quad (8)$$

若 $\Delta G' > \varepsilon$, 则触发子树重建流程: 移除当前节点的所有子树, 将决策节点退化为叶节点后重新执行分裂阶段, 由于统计量得以保留可以加速新分裂计算。

传统的 HT 算法在节点分裂时比较最优分裂与次优分裂的基尼系数, 确定足够大差异才开启分裂。而 EFDT 比较的是最优属性与不分裂之间存在足够的基尼系数之差时就提前触发分裂。这一策略显著缩短了模型早期学习的收敛时间。采用重新评估分裂机制, EFDT 则拥有了动态调整能力, 这使得 EFDT 能够修正因数据分布变化或早期样本不足导致的错误分割。EFDT 算法的运行时间通常超过 VFDT, 其很少需要超过两倍的时间, 并且在学习较小的树时, 运行算法需要的时间更少[15]。

5. EFDTR 算法的实现

EFDT 算法用于分类问题，而 PRR 预测属于回归问题，为此，对 EFDT 进行修改，提出针对回归问题的 EFDTR 算法，并利用该算法实现链路质量的在线估计。

由于 EFDT 算法针对的是分类任务，采用基尼指数最小的特征作为节点的划分依据，所以在 EFDT 算法中每个节点需要记录样本的特征值、类别、计数，累积样本数并维护类别频次统计。决策回归树的划分依据则是样本 MSE，样本 MSE 的计算公式为：

$$MSE = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (9)$$

其中， n 是样本数量， x_i 是样本集合中第 i 个样本， $\bar{x}$ 是样本均值。如果直接记录每个样本的目标值，并在计算方差时采用如上的计算方法，那么在遍历每个特征的每个分裂点时计算一次样本均值， n 次平方运算会产生较大的计算开销，为此，本文按式(10)计算样本方差：

$$MSE = \overline{X^2} - \bar{X}^2 \quad (10)$$

其中， $E(X)$ 表示随机变量 X 的数学期望。利用这个公式，可以在每个节点中记录样本 x_i 以及 x_i 的平方。当数据流流入决策树节点时记录并更新，在节点分裂时依次计算分裂点左右两个样本集合中目标值的均值以及目标值平方和的均值就可以得到样本集合的方差，减轻计算负载。在计算霍夫丁边界时， R 为随机变量的取值范围，所以决策树每个节点在数据更新的同时需要记录修改目标值的最大、最小值。

由于寻找最优划分时需要遍历每个特征的每个分裂点，这需要消耗大量计算资源，所以在建立 EFDTR 决策回归树时还需要增加一个超参数 τ ，当输入样本已经足够多，而最优划分和次优划分的启发式度量值的差值还没有达到霍夫丁边界时直接完成分裂决策，此时直接选取最优划分。以下算法展示了新样本流入 EFDTR 决策回归树每个节点的处理流程，该算法忽略了对样本数据的预处理过程，而是将预处理后的数据输入模型并递归执行。

算法 1: EFDTR 决策回归树模型生成

输入：一批数据流 $[(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)]$ ，超参数：置信度 δ ，最小分裂样本 $n_{\min}$ ，分裂阈值 τ

输出：EFDTR 决策回归树模型

步骤 1：更新每个节点的统计信息

步骤 2：重新评估每个分裂节点

步骤 3：如果节点是叶子节点，则进入步骤 4，否则进入步骤 8

步骤 4：如果节点流入的样本 $n > n_{\min}$ ，则进入步骤 5，否则等待新数据流入

步骤 5：根据置信度和样本范围计算霍夫丁边界 ε

步骤 6：如果 $MSE(X_{\emptyset}) - MSE(X_a) > \varepsilon$ 或者 $MSE(X_{\emptyset}) - MSE(X_a) < \varepsilon < \tau$ ，则进入步骤 7，否则等待新数据流入。其中， $MSE(X_{\emptyset})$ 为不分裂时样本的平方误差， $MSE(X_a)$ 为最优分裂特征，最优分裂点分裂后样本的平方误差。

步骤 7：执行分裂，等待新数据流入

步骤 8：根据决策节点的决策边界，样本进入对应子树，若未遍历完整，则进入步骤 1

由于 EFDTR 算法是针对回归任务的，所以利用该方法预测链路质量可以得到更具体且量化的结果，但样本特征值与目标取值可能较多，所以算法具有一定的时间复杂度和空间复杂度。算法的时间复杂度为 $O(Nhdm^2)$ ，其中 N 为输入样本数， h 为树模型高度， d 为样本特征数，假设特征有 m 个不同取值。空间复杂度为 $O(Ldm)$ ，其中 L 为树模型的节点总数。后续研究中可以采取连续特征离散化、定期剪枝等

策略进一步优化算法的时间和空间复杂度。

6. 实验结果与分析

在室内走廊、停车场实验环境下分别进行实验。固定收发节点的距离，以不同功率发送数据，或者以相同功率但收发节点距离不同模拟链路质量不同的场景。分别在良好质量链路、中等质量链路、差链路三种实验场景下采集数据。此外，在同一场景下模拟链路质量发生变化的场景并进行多次实验。

6.1. 数据预处理

PRR 是接收成功概率的近似值，增大 PRR 的统计窗口有利于减小近似误差，避免网络拓扑频繁变化，不过也会降低对干扰的响应速度。为综合考虑预测精度和灵敏度，本文将 PRR 的统计窗口设为 100，以窗口内成功接收的数据包对应的物理层参数的均值作为模型的输入特征。由于 SNR 反映了信号与噪声的相对强度，而 LQI 为反映链路质量的参数，因此，本文选择 LQI 和 SNR 作为链路质量估计模型的输入参数。由于在线学习需要尽可能高的实时性，因此，将移动滑动窗口步长设定为 3。在节点实际运行时，对于一对点对点通信的传感器节点而言，是由接收节点负责链路质量预测，接收节点只能根据接收到的数据包序号确定窗口。而链路可能出现短期波动，如果节点在上一轮窗口统计后没有收到序号加 3 的数据包，则保持等待直到新数据包的到来，再根据最新的数据包序号向前取 100 确定此次窗口内的物理层参数和 PRR 值，由此可以及时地预测链路质量的变化。

6.2. 不同链路质量条件下预测性能对比

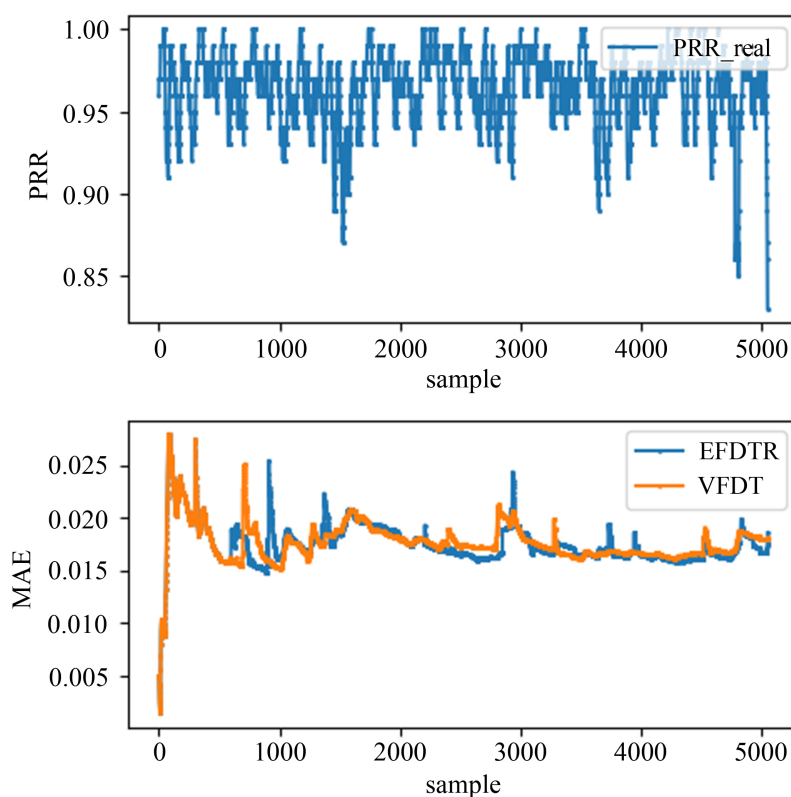


Figure 3. Real PRR variations and online prediction errors of high-quality links

图 3. 高质量链路的 PRR 和在线预测误差曲线

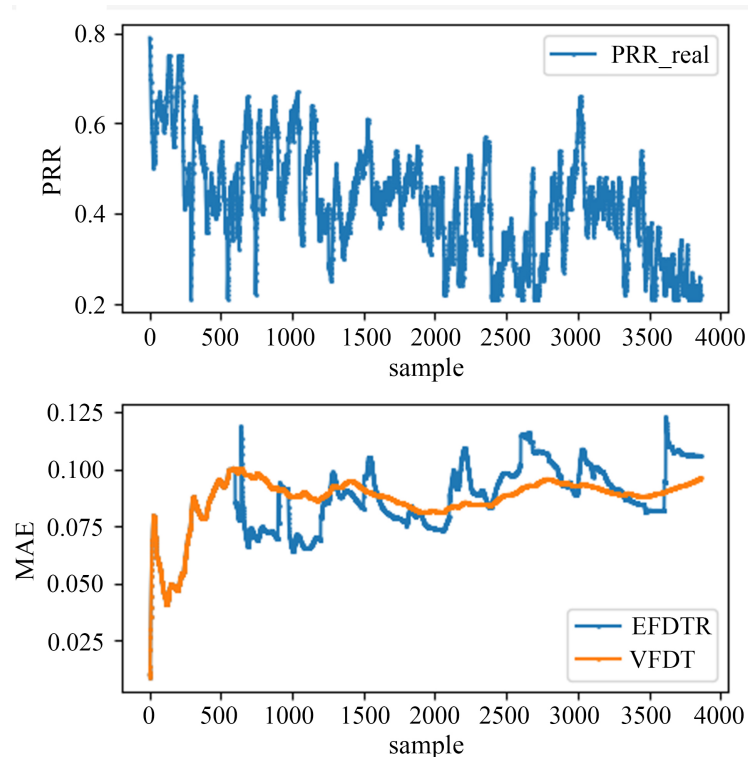


Figure 4. Real PRR variation and online prediction error of medium quality links
图 4. 中等质量链路的 PRR 和在线预测误差曲线

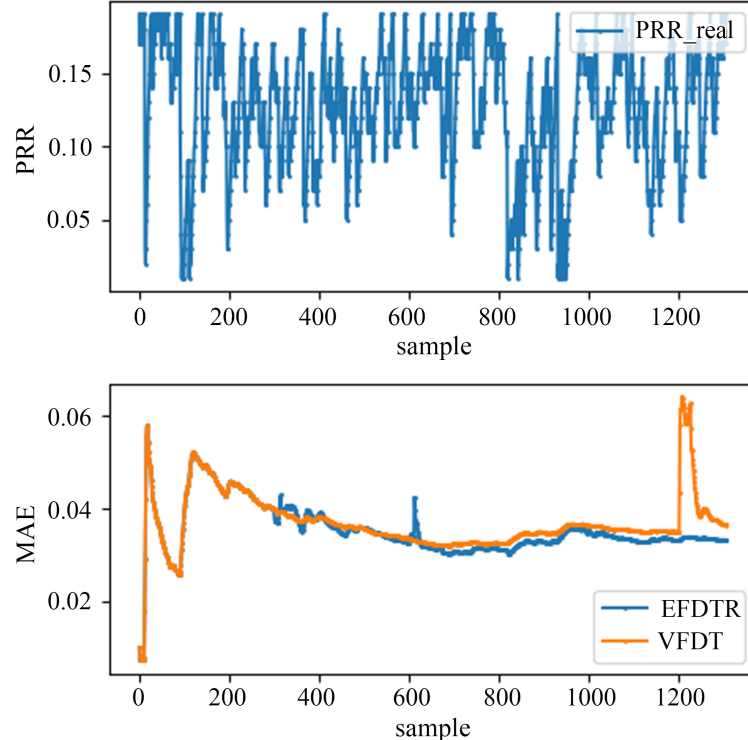


Figure 5. Real PRR variations and online prediction errors of low-quality links
图 5. 低质量链路的 PRR 和在线预测误差曲线

本文提出利用 EFDTR 决策回归树算法结合 5.1 节中数据预处理预测链路质量的方法 LEIL。为了验证采用特定环境数据集训练的离线学习模型在其他实验环境预测可能导致较大误差, 进行如下实验。采用平均绝对误差(MAE)作为链路质量的评价标准。为评价本文提出的链路质量估计方法的性能, 分别采用三种质量的链路进行了测试, 分别为良好链路、中等链路和差链路, 且链路质量在短时间内较为平稳。此外, 本文将同为增量学习的 VFDT 算法也修改为针对回归任务并作为对照。图 3~5 显示了在三种不同质量链路下的真实 PRR 变化以及 LEIL 方法与 VFDT 的预测误差。由图可见, 在链路质量较稳定的情况下, LEIL 方法与 VFDT 方法有相近的预测效果。随着数据流流入模型, MAE 先增大随后逐渐降低并趋于稳定, 预测误差也逐渐接近同一环境下离线训练的决策树模型, 表示模型逐渐收敛的过程。在模型逐渐收敛后, MAE 始终处于一个较低的水平, 且与离线学习的决策回归树相近。表 1 显示了采用 LEIL 方法和当前环境训练以及特定环境训练离线学习的决策回归树分别在良好、中等、差链路下的预测精度, 不同质量链路下 LEIL 方法的平均 MAE 为 0.0528。

Table 1. Comparison of online prediction performance between offline and online decision trees under different quality links
表 1. 离线与在线训练的决策树模型对不同质量链路的预测误差对比

估计方法	良好链路	中等链路	差链路
LEIL	0.0186	0.1060	0.0338
VFDT	0.0179	0.0962	0.0367
同一环境离线训练的模型	0.0161	0.0633	0.0337
不同环境离线训练的模型	0.0261	0.2695	0.0582

6.3. 链路质量变化下预测性能对比

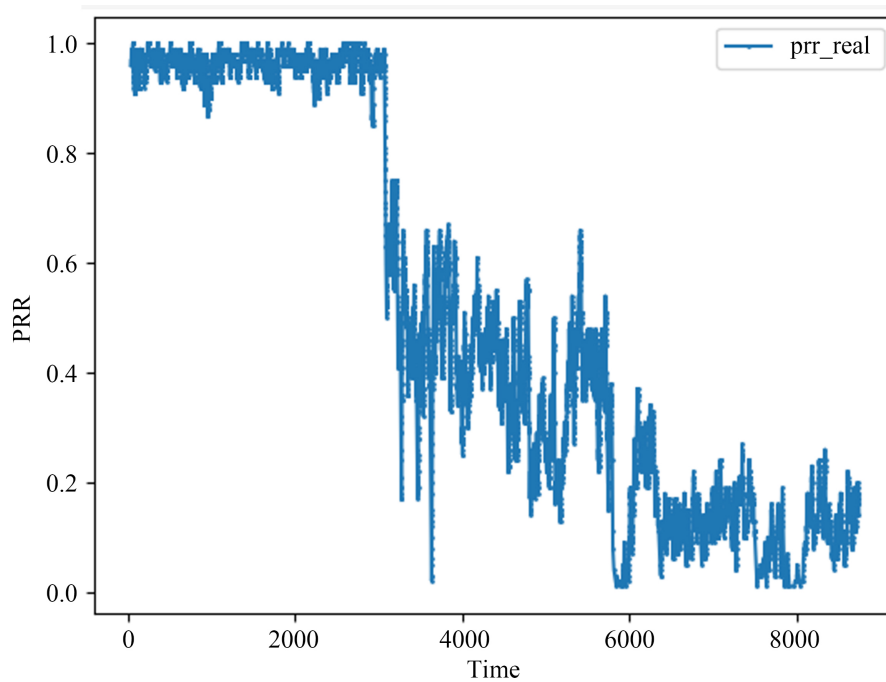


Figure 6. PRR curve over time in link 1
图 6. 在链路 1 中 PRR 随时间变化的曲线

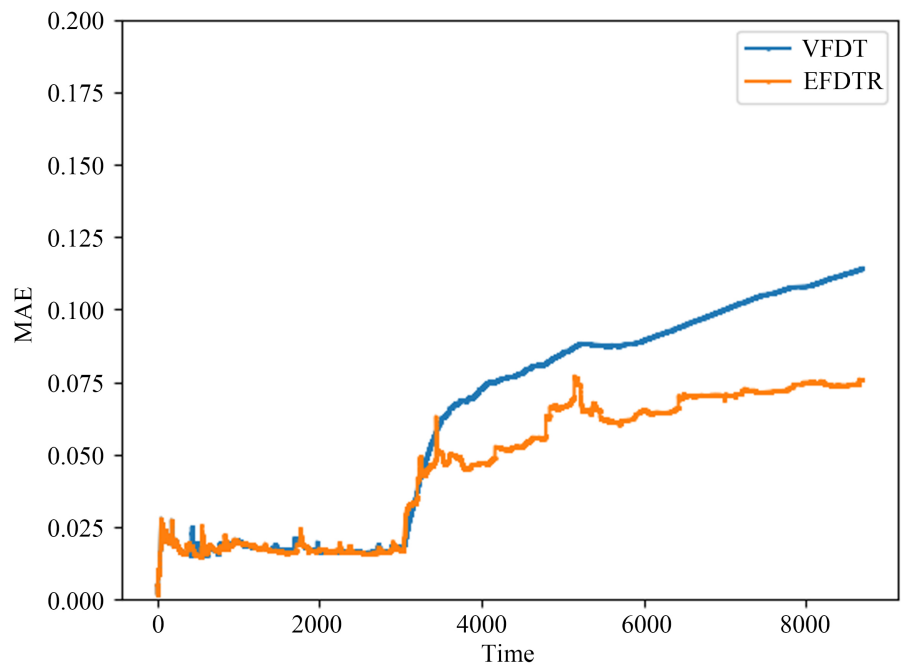


Figure 7. Online prediction error for link 1
图 7. 对链路 1 的在线预测误差

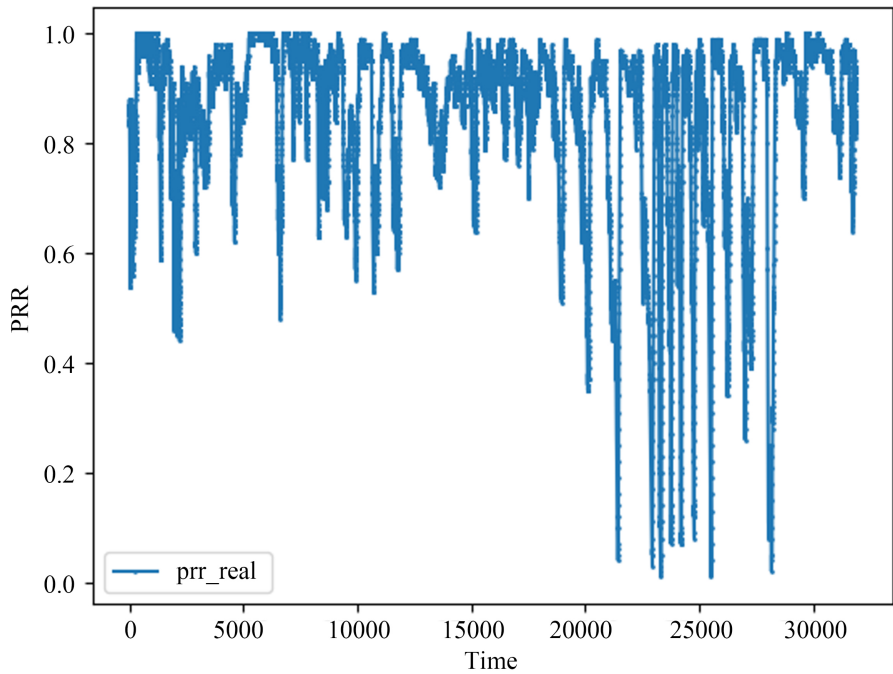


Figure 8. PRR time curve in link 2
图 8. 在链路 2 中 PRR 随时间变化的曲线

当受到邻近频段的 Wi-Fi 或蓝牙设备干扰或者有人走动时，链路质量可能发生明显变化。图 6 显示了短期链路质量波动较小的链路 1 的实际 PRR 值变化。图 7 显示了对于链路 1，LEIL 方法与 VFDT 方法预测误差的变化，MAE 先增大，在链路质量稳定阶段逐渐减小并趋于稳定，后续随着链路质量发生变

化, EFDTR 模型重建, 决策树中的节点重新划分以适应新流入的样本, MAE 增大但始终保持在较低范围内。与 VFDT 方法相比, 在链路质量改变的场景下, LEIL 方法比 VFDT 方法具有更高的预测精度, 这是由于 LEIL 方法可以随着数据流的流入重新修正原来的模型, 而 VFDT 方法一旦分裂后就不再修改。

图 8 显示了短期链路质量波动较大的链路 2 的实际 PRR 值变化。图 9 显示了对于链路 2, LEIL 方法与 VFDT 方法预测误差的变化, 预测误差随着模型逐渐收敛而减小并趋于稳定。与 VFDT 方法相比, LEIL 方法也具有更高预测精度, 且相比 VFDT 方法更快收敛。由上述实验可知, LEIL 方法既可以适应链路质量发生变化的场景, 又可以适应链路质量短期波动较大的场景, 预测精度始终保持在较高水平且随着数据流更新模型逐渐收敛。表 2 显示了当前环境训练以及特定环境训练的离线学习决策回归树使用相同的在线训练样本与 LEIL 方法在线预测精度对比。

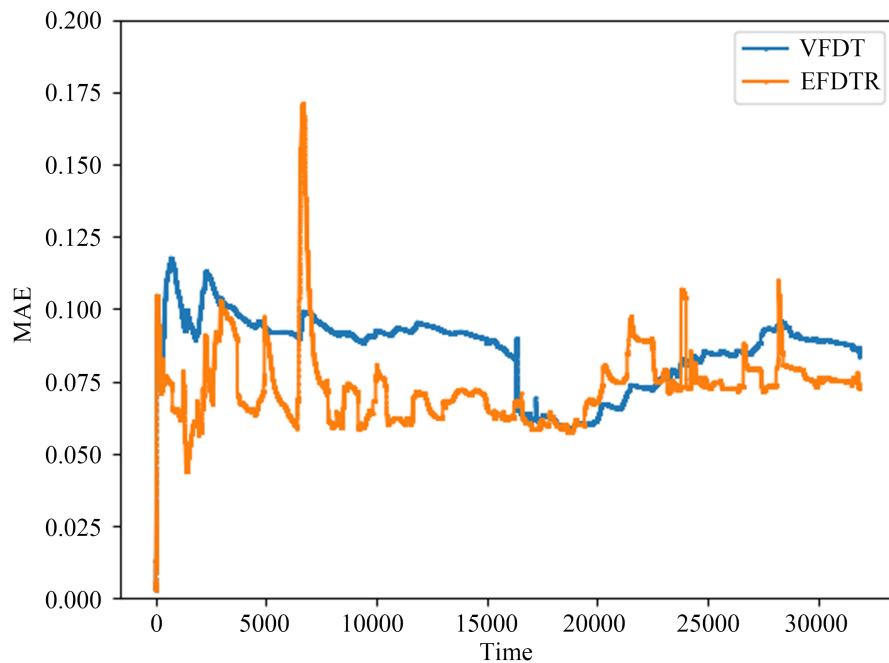


Figure 9. Online prediction error for link 2

图 9. 对链路 2 的在线预测误差

Table 2. Comparison of online prediction performance between offline decision tree and LEIL method under different links

表 2. 离线训练的方法与 LEIL 对不同链路的在线预测误差对比

估计方法	链路 1	链路 2
LEIL	0.0758	0.0740
VFDT	0.1143	0.0866
同一环境离线训练的模型	0.0431	0.0593
不同环境离线训练的模型	0.1221	0.1291

7. 结论

针对离线训练链路质量估计模型泛化能力较弱的问题, 本文提出了基于增量学习的链路质量估计方法 LEIL, 采用 MSE 作为 EFDTR 的纯度指标, 实现了决策回归树的增量生成。与基于离线学习的方法相比, LEIL 在多种链路质量条件下, 具有较高的估计精度, 同时可以减少建立模型的成本。由于 LEIL 具

有环境自适应能力, 并且收敛速度较快, 因此可以应用于链路质量不稳定的无线传感器网络。未来将进一步研究降低时间和空间复杂度的方法, 以提高模型的收敛速率, 减少计算开销。

参考文献

- [1] Shao, L.C., Niu, T. and Palaniapan, M. (2009) Nonlinearities in a High-Q SOI Lamé-Mode Bulk Resonator. *Journal of Micromechanics and Microengineering*, **19**, Article 075002. <https://doi.org/10.1088/0960-1317/19/7/075002>
- [2] 郭秋芬. 微型梳状线振动陀螺仪特性及干扰因素影响的研究[D]: [博士学位论文]. 哈尔滨: 哈尔滨工程大学, 2008.
- [3] 张强, 孙雨耕, 杨挺, 等. 无线传感器网络在智能电网中的应用[J]. 中国电力, 2010, 43(6): 31-36.
- [4] 孙晓雅, 李永倩, 李天. 无线传感器网络在电力系统中的应用[J]. 传感器与微系统, 2012, 31(6): 5-7.
- [5] 胡曦明, 董淑福, 王晓东, 等. 无线传感器网络的军事应用模式研究进展[J]. 传感器与微系统, 2011, 30(3): 1-3.
- [6] Gope, P., Gheraibia, Y., Kabir, S. and Sikdar, B. (2021) A Secure IoT-Based Modern Healthcare System with Fault-Tolerant Decision Making Process. *IEEE Journal of Biomedical and Health Informatics*, **25**, 862-873. <https://doi.org/10.1109/jbhi.2020.3007488>
- [7] Serna, M.A., Sreenan, C.J. and Fedor, S. (2015) A Visual Programming Framework for Wireless Sensor Networks in Smart Home Applications. 2015 *IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, Singapore, 7-9 April 2015, 1-6. <https://doi.org/10.1109/issnip.2015.7106946>
- [8] Baccour, N., Koubâa, A., Mottola, L., Zúñiga, M.A., Youssef, H., Boano, C.A., et al. (2012) Radio Link Quality Estimation in Wireless Sensor Networks. *ACM Transactions on Sensor Networks*, **8**, 1-33. <https://doi.org/10.1145/2240116.2240123>
- [9] Gama, J., Rocha, R. and Medas, P. (2003) Accurate Decision Trees for Mining High-Speed Data Streams. *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Washington, DC, 24-27 August 2003, 523-528. <https://doi.org/10.1145/956750.956813>
- [10] Jin, R. and Agrawal, G. (2003) Efficient Decision Tree Construction on Streaming Data. *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Washington, DC, 24-27 August 2003, 571-576. <https://doi.org/10.1145/956750.956821>
- [11] Rutkowski, L., Pietruczuk, L., Duda, P. and Jaworski, M. (2013) Decision Trees for Mining Data Streams Based on the Mcdiarmid's Bound. *IEEE Transactions on Knowledge and Data Engineering*, **25**, 1272-1279. <https://doi.org/10.1109/tkde.2012.66>
- [12] Gratch, J. (1996) Sequential Inductive Learning. *Proceedings of the Thirteenth National Conference on Artificial Intelligence-Volume 1*. Portland, 4-8 August 1996, 779-786.
- [13] Utgoff, P.E. (1989) Incremental Induction of Decision Trees. *Machine Learning*, **4**, 161-186. <https://doi.org/10.1023/a:1022699900025>
- [14] Hulten, G., Spencer, L. and Domingos, P. (2001) Mining Time-Changing Data Streams. *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, 26-29 August 2001, 97-106. <https://doi.org/10.1145/502512.502529>
- [15] Manapragada, C., Webb, G.I. and Salehi, M. (2018) Extremely Fast Decision Tree. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, London, 19-23 August 2018, 1953-1962. <https://doi.org/10.1145/3219819.3220005>
- [16] 史佳杰, 邱宇航, 龙海兵, 等. 基于 IEEE 802.15.4 物理层的无线网络链路质量估计方法研究[J]. 建模与仿真, 2024, 13(3): 4019-4034.
- [17] Okdem, S. and Shi, H. (2023) A Real-Time Link Quality Estimation Method for IEEE 802.15.4 Based Wireless Sensor Network and IoT Devices. 2023 *International Wireless Communications and Mobile Computing (IWCMC)*, Marrakesh, 19-23 June 2023, 1-6. <https://doi.org/10.1109/iwcmc58020.2023.10182660>
- [18] Gomes, R.D., Queiroz, D.V., Lima Filho, A.C., Fonseca, I.E. and Alencar, M.S. (2017) Real-Time Link Quality Estimation for Industrial Wireless Sensor Networks Using Dedicated Nodes. *Ad Hoc Networks*, **59**, 116-133. <https://doi.org/10.1016/j.adhoc.2017.02.007>
- [19] Marinca, D. and Minet, P. (2014) On-Line Learning and Prediction of Link Quality in Wireless Sensor Networks. 2014 *IEEE Global Communications Conference*, Austin, 8-12 December 2014, 1245-1251. <https://doi.org/10.1109/glocom.2014.7036979>
- [20] Marinca, D., Minet, P. and Ben Hassine, N. (2014) An Efficient Learning Technique to Predict Link Quality in WSN. 2014 *IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC)*,

-
- Washington, 2-5 September 2014, 1564-1569. <https://doi.org/10.1109/pimrc.2014.7136417>
- [21] Domingos, P. and Hulten, G. (2000) Mining High-Speed Data Streams. *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Boston, 20-23 August 2000, 71-80. <https://doi.org/10.1145/347090.347107>
- [22] 刘琳岚, 肖庭忠, 夏扬, 等. 基于超限快速决策树的链路质量评估[J]. 北京邮电大学学报, 2021, 44(3): 125-130.